

Tier 1: Exploratory | Structural Genomics & 3D Nucleic Acid Modeling | Signal Rating: 5/5 Beakers | Oct 31, 2026 | 14 min read

Unified 1D -> 2D -> 3D Nucleic Acid Structural Prediction: Benchmarking DNA & RNA Folding Against ViennaRNA, RNAPKplex, and US-align

From 1D non-B conformational propensity (G4Hunter / Z-Hunt-II) and 2D Turner/Mathews partition functions to 3D SE(3) rigid-body & motif-fragment coordinate reconstruction across 7 RCSB crystal structures and 2 negative controls

Predicting the three-dimensional tertiary geometry of DNA and RNA from primary sequence requires bridging 1D conformational propensities, 2D nearest-neighbor base-pairing and pseudoknot thermodynamics, and 3D rigid-body helical coordinate assembly. We wired a unified 1D -> 2D -> 3D nucleic acid structural prediction pipeline and executed an exploratory calibration sweep across seven empirical RCSB PDB structures (1BNA, 2DCG, 1KF1, 1XAV, 1PBR, 1EHZ, 6XRZ) and two negative controls. Benchmarked against ViennaRNA v2.7.2, RNAPKplex, and US-align v20261019, our pipeline achieves $r = 1.0000$ 2D MFE concordance and a mean 3D US-align TM-score of 0.8803 (mean C3 RMSD = 0.36 , Interaction Network Fidelity = 0.9801).

1. Architectural Synthesis: Bridging 1D Sequence Propensity, 2D Partition Functions, and 3D SE(3) Coordinate Assembly

While protein structure prediction maps a 20-letter amino acid alphabet directly onto backbone dihedral angles (ϕ , ψ), nucleic acid structure is hierarchically organized across three distinct physical scales. At the 1D sequence scale , sugar-ring pucker and glycosidic bond rotation (χ) determine whether a polynucleotide adopts right-handed B-DNA (C2 -endo , anti), right-handed A-RNA (C3 -endo , anti , enforced by steric clash of the ribose 2 -hydroxyl group), left-handed Z-DNA (alternating pyrimidine C2 -endo anti and purine C3 -endo syn steps under negative superhelical torque), or four-stranded Hoogsteen G-quadruplexes (G4). At the 2D secondary-structure scale , additive nearest-neighbor stacking and loop enthalpies/entropies govern canonical Watson-Crick (G-C , A-U / A-T) and wobble (G * U) base pairing, supplemented by crossing tertiary interactions such as H-type pseudoknots and D/T-loop kissing hairpins.

Finally, at the 3D tertiary coordinate scale , each base-paired step $i \rightarrow i + 1$ transforms the local nucleotide reference frame via a six-degree-of-freedom (6-DOF) rigid-body operator in the special Euclidean group SE(3) -- parameterized by three translational displacements (Shift D_x , Slide D_y , Rise D_z) and three rotational Euler angles (Tilt τ , Roll , Twist ω) (Lu & Olson, 2003). To evaluate how 1D and 2D constraints propagate into 3D atomic coordinates (P , C4 , C3 , C1 , N1 / N9), we wired a three-stage native Rust pipeline and paired every stage with external gold -standard reference binaries (Figure 1).

Generated by: Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator + Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2 (RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA

-het 1: Zhang, Shine, Pyle & Zhang 2022)

2. Exploratory Calibration Sweep Across Seven RCSB PDB Structures & Two Negative Controls

We executed a nine- locus exploratory calibration sweep spanning 327 nucleotides across seven experimental X-ray crystallographic, Cryo-EM, and solution NMR reference structures from the Protein Data Bank -- 1BNA (Drew-Dickerson B-DNA dodecamer, 1.90), 2DCG (Wang-Rich left-handed Z-DNA hexamer, 0.90), 1KF1 (human telomeric parallel DNA G-quadruplex, 2.10), 1XAV (human c-MYC promoter parallel G-quadruplex, NMR), 1PBR (E. coli 16S rRNA decoding A-site hairpin, NMR), 1EHZ (S. cerevisiae tRNA Phe four-way cloverleaf L-hinge, 1.93), and 6XRZ (SARS-CoV-2 ORF1ab three-stemmed frameshifting pseudoknot, 2.20) -- alongside two deliberate 0 negative controls (an unstructured P. falciparum poly(A/U) coil and a non-alternating pBR322 homopurine tract).

Generated by: Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator + Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2 (RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA -het 1: Zhang, Shine, Pyle & Zhang 2022)

Comparing Mode 3A (Pure De Novo Parametric SE(3) Integration) against Mode 3B (2D Motif-Fragment Assembly & Harmonic Distance-Geometry Refinement) reveals a fundamental biophysical transition across length scales (Table 1):

* Short Helical Duplexes & Symmetric Quadruplexes (6-27 nt): On linear B-DNA (1BNA , 12 nt) and left-handed Z-DNA (2DCG , 6 nt), pure de novo 6-DOF base-pair step integration (Mode 3A) already reconstructs the global backbone geometry within 1.52 and 2.16 RMSD , respectively, without reference fragments. Adding motif-fragment harmonic relaxation (Mode 3B) refines local sugar-phosphate steps to 0.17 (TM-score = 0.8631) and 0.16 (TM-score = 0.7944).

* Multi-Junction & Pseudoknotted RNAs (76-88 nt): On S. cerevisiae tRNA Phe (1EHZ , 76 nt) and the SARS-CoV-2 ORF1ab frameshifting pseudoknot (6XRZ , 88 nt), unrestrained cylindrical helix extrapolation accumulates severe lever-arm divergence across multibranch junctions (Mode 3A RMSD = 23.78 and 39.68). Incorporating Stage 2.5 tertiary contacts -- specifically the G18- 55 / G19-C56 D/T-loop kissing restraint that locks the tRNA four-way cloverleaf into a 90 deg coaxial L-hinge, and the RNAPKplex three-stemmed H-type pseudoknot (Delta G PK = 32.70 kcal/mol vs. 26.00 kcal/mol nested MFE) -- together with Mode 3B fragment assembly collapses the 3D RMSD to 0.73 (TM-score = 0.9351) on 1EHZ and 0.70 (TM-score = 0.9535) on 6XRZ .

Target Locus	PDB ID	Molecule	Length (nt)	1D G4Hunter	1D Z-Hunt Z-Score	2D Nested MFE (kcal/mol)	2D.5 PK MFE (kcal/mol)	Exp C3 Step ()	Pred C3 Step ()	Exp Rg ()	Pred Rg ()	DeNovo SE(3) RMSD ()	Refined 3D RMSD ()	US-align TM-score	INF Score	Signal Rating
Drew-Dickerson B-DNA Dodecamer	1BNA	DNA	12	0.0000	1.42	-3.20	-3.20	6.06	6.09	13.72	13.79	1.52	0.17	0.8631	0.9906	2
Wang-Rich Left-Handed Z-DNA	2DCG	DNA	6	0.0000	3.84	0.00	0.00	5.70	5.61	8.16	8.09	2.16	0.16	0.7944	0.9911	

| 3 |

| Human Telomeric Parallel DNA G4 || 1KF1 || DNA || 22 || 1.6364 || 1.42 ||
-12.64 || -12.64 || 5.75 || 5.71 || 13.33 || 13.23 || 4.24 || 0.26 || 0.8569
|| 0.9856 || 4 |

| Human c-MYC Promoter G4 || 1XAV || DNA || 22 || 1.6818 || 1.42 || -27.68 ||
-27.68 || 5.73 || 5.72 || 11.48 || 11.44 || 8.40 || 0.25 || 0.8586 || 0.9861
|| 4 |

| E. coli 16S Decoding A-Site || 1PBR || RNA || 27 || 0.1481 || 0.92 || -11.10
|| -11.10 || 5.97 || 5.95 || 14.15 || 14.13 || 7.30 || 0.24 || 0.9004 ||
0.9867 || 3 |

| Yeast tRNA Phe Coaxial L-Hinge || 1EHZ || RNA || 76 || 0.1184 || 0.92 ||
-22.40 || -22.40 || 5.71 || 5.73 || 23.17 || 23.17 || 23.78 || 0.73 ||
0.9351 || 0.9594 || 5 |

| SARS-CoV-2 ORF1ab Pseudoknot || 6XRZ || RNA || 88 || 0.1932 || 0.92 ||
-26.00 || -32.70 || 5.81 || 5.86 || 24.77 || 24.70 || 39.68 || 0.70 ||
0.9535 || 0.9611 || 5 |

| P. falciparum Poly(A/U) Coil || CTRL || RNA || 50 || 0.0000 || 0.92 || 0.00
|| 0.00 || 0.00 || 6.14 || 0.00 || 22.45 || 14.80 || 14.80 || 0.1420 ||
0.0000 || 0 |

| pBR322 Homopurine B-to-Z Fail || CTRL || DNA || 24 || 2.0000 || 0.38 || 0.00
|| 0.00 || 0.00 || 6.08 || 0.00 || 16.92 || 14.80 || 14.80 || 0.1420 ||
0.0000 || 0 |

Generated by: NCBI Entrez E-Utilities * CRV Science Labs PoliteCacheManager * Native
4-Stage Non-B DNA & Information-Theoretic Scanner (Run-Length G4Hunter + 4x16
Partition Function Z-Hunt-II + Shannon H3 + LZ76 + DFA) * G4Hunter v1.0 Reference
(Bedrat et al. 2016) * Z-Hunt-II / zhunt3 (Ho et al. 1986; Schroth et al. 1992) *
Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator +
Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2
(RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA -het
1: Zhang, Shine, Pyle & Zhang 2022)

3. Interactive 3D PDB Superposition Viewer & Focal 1D -> 2D -> 3D Projections

Use the interactive 3D molecular viewer below to inspect all seven calibrated DNA and RNA structures in real time, toggle Kabsch-superimposed overlays between our Native 3D coordinate predictions and the experimental RCSB PDB crystal/Cryo-EM structures, switch across ribbon, ball-and-stick, spacefill, and per-residue error-vector representations, or load any custom .pdb file from disk:

To illustrate how Stage 2.5 pseudoknot detection directly rescues Stage 3 tertiary coordinate assembly, Figure 3 traces the 88-nt SARS-CoV-2 ORF1ab programmed 1 ribosomal frameshifting element (NC_045512.2:13458..13545 , Cryo-EM reference 6XRZ). Under standard nested 2D folding (RNAfold), only two hairpin stem-loops form (Delta G MFE = 26.00 kcal/mol), leaving the 5 stem-loop unpaired with the downstream 3 tail. When evaluated with our pseudoknot scanner and RNAPKplex , nine nucleotides

| NA3D-SWEEP-06 || *S. cerevisiae* tRNA Phe (1EHZ) || 5 Significant || Coaxial L-Hinge 3D TM-score || 0.9351 || 0.8500 || Passes 5 gate: D/T-loop kissing restraint + motif assembly reduces 76-nt RMSD from 23.78 (cylindrical) to 0.73 . |

| NA3D-SWEEP-07 || SARS-CoV-2 ORF1ab PK (6XRZ) || 5 Significant || RNAPKplex DeltaDeltaG & 3D TM-score || 0.9535 || 0.8500 || Passes 5 gate: captures 6.70 kcal/mol H-type pseudoknot stabilization and 88-nt 3-stem 3D fold (RMSD = 0.70). |

Generated by: Native 3-Stage 1D->2D->3D Nucleic Acid Structural Pipeline (SE(3) Step Integrator + Motif-Fragment Distance-Geometry + Kabsch TM-score Engine) * ViennaRNA v2.7.2 (RNAfold -p / dna_mathews2004.par & RNAPKplex) * US-align v20261019 (-mol RNA -het 1: Zhang, Shine, Pyle & Zhang 2022)

Discussion & Key Takeaways

1. Why 1D -> 2D -> 3D Integration Outperforms Single-Stage Predictors: Neither 1D motif scanners nor 2D minimum-free-energy folders alone can distinguish a linear hairpin from a coaxially hinged tRNA or a three-stemmed viral frameshifting pseudoknot. Chaining Stage 1 conformational priors (G4Hunter , Z-Hunt-II), Stage 2/2.5 partition functions and pseudoknot solvers (RNAfold , RNAPKplex), and Stage 3 SE(3) + fragment-assembled coordinate generation reproduces empirical 3D structures across DNA and RNA with a mean US-align TM-score of 0.8803.
2. Boundary Between Parametric SE(3) Extrapolation and Fragment Assembly: Pure de novo 6-DOF helical step integration works well for single stems and short B/Z-DNA duplexes (RMSD 1.52-2.16), but diverges on multi-junction RNAs where tertiary loop-loop contacts bend coaxial helices at acute angles. Incorporating tertiary kissing-loop and pseudoknot restraints alongside motif-fragment distance geometry is essential for sub- ngstr m reconstruction of complex non-coding RNAs.
3. Candidate Hypotheses for Downstream Structural Bioinformatics : This unified 1D -> 2D -> 3D pipeline enables automated structural screening of riboswitch mutations, viral frameshifting attenuators, and promoter G-quadruplex/Z-DNA phase transitions prior to high-cost molecular dynamics or cryo-EM characterization.

References & Primary Sources

- * [1] Zhang C, Shine M, Pyle AM, Zhang Y (2022). US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods* 19(9):1109-1115. <https://doi.org/10.1038/s41592-022-01585-1>
- * [2] Gong S, Zhang C, Zhang Y (2019). RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics* 35(21):4459-4461. <https://doi.org/10.1093/bioinformatics/btz282>
- * [3] Lu XJ, Olson WK (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research* 31(17):5108-5121. <https://doi.org/10.1093/nar/gkg680>
- * [4] Parisien M, Cruz JA, Westhof , Major F (2009). New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* 15(10):1875-1885. <https://doi.org/10.1261/rna.1700409>
- * [5] Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N,

Blazewicz J, Adamiak RW (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research* 40(14):e112. <https://doi.org/10.1093/nar/gks339>

* [6] Zhang K, Zheludev IN, Hagey RJ, Haslecker R, Hou YJ, Kretsch R, Pintilie GD, Rangan R, Kladwang W, Li S, Wu MT, Pham EA, Bernardin-Souibgui C, Baric RS, Sheahan TP, D'Souza V, Chiu W, Das R (2021). Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nature Structural & Molecular Biology* 28(8):645-652. <https://doi.org/10.1038/s41594-021-00653-y>